



**Big Data & Analytics**

# Privacy in the era of Big Data and Open Government

Kathryn Zeidenstein  
InfoSphere Guardium Evangelist  
[krzeide@us.ibm.com](mailto:krzeide@us.ibm.com)



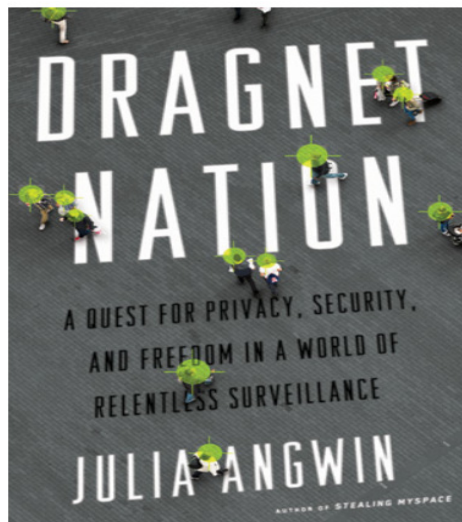
# Agenda

- Data privacy and security hit parade
- Embracing new technology in government Initiatives
- Data privacy and security considerations for big data
- IBM's approach to privacy in big data initiatives





## Trust and Accuracy



### **What Do Data Brokers Know About Me?**

I tried to find out. Some of their information is frighteningly accurate—and some of it laughably wrong.



60 MINUTES OVERTIME

**SHOCKED TO LEARN HOW DATA BROKERS ARE WATCHING YOU?**



# BIG DATA



**Kashmir Hill**  
Forbes Staff

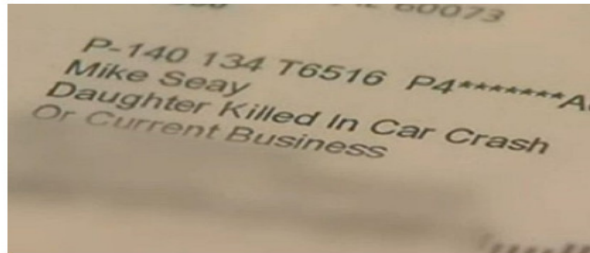
FOLLOW

Welcome to The Not-So Private Parts where technology & privacy collide  
[full bio](#) →



## OfficeMax Blames Data Broker For 'Daughter Killed in Car Crash' Letter

1 comments, 1 called-out + Comment Now + Follow Comments



Yikes. [OfficeMax](#) [OMX.NA%<sup>1</sup>](#) is getting a lot of press these days and it has little to do with its office supply deals. It was a little too specific in targeting one of its customers, sending a mailing to a Chicago man named Mike Seay addressed to “Daughter Killed in Car Crash or Current Business.” Seay’s teenage daughter had been killed in a car accident a year earlier, as reported by [NBC News](#)

IBM Commits \$1.2B to Expand Global Cloud Footprint

Enabling Innovation With Cloud



[Blog: Cloud Computing G](#)  
[Infographic: Network of Cl](#)  
[SoftLayer Trial: Give IBM C](#)

Tweets from IBM Cloud

IBMCloud @IBMcloud  
IBM is up in the #cloud, accord



↑ CONFERENCES AND MORE

# The breaches keep coming...



NOV '13

Personal info and credit card numbers of **tens of millions of Target customers** were stolen during prime shopping season from November through December.



APR '14

A federal judge in California denied LinkedIn's motion to throw out a putative class action lawsuit over a 2012 breach that resulted in **6.5 million stolen passwords**, Plaintiff claims she was swayed by **misleading labeling on privacy policy**.



SEPT '13

**Vodafone Germany**: personal information on more than two million mobile phone customers has been stolen, extracted from an internal databases by an insider

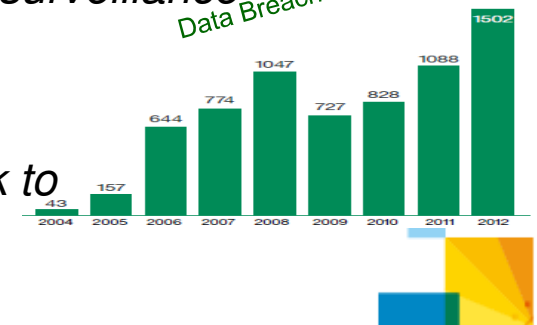


The Chinese government attack on Google servers in 2010 accessed a sensitive database with years of info on US government surveillance targets



Jan '14

An IT consultant at **Korea Credit Bureau** copied names, SSNs and credit card details of millions onto a USB stick to sell to a marketing firm. Data was not encrypted.



# Security incidents by country and sector

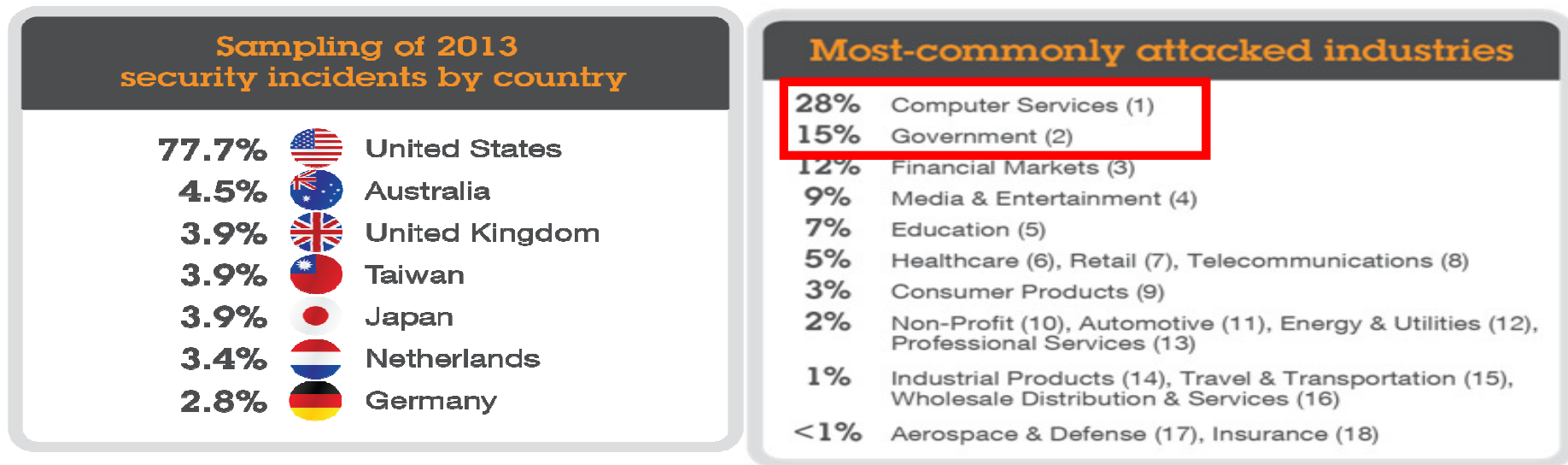


Figure 3. Sampling of 2013 security incidents by country

Source: IBM X-Force® Research and Development

**.5 BILLION RECORDS OF PII WERE LEAKED IN 2013!**



# Lessons from the Edward Snowden/NSA Data Breach



**Snowden case: How low-level insider cracked NSA**

Byron Acochido and Peter Eisler, USA TODAY 7:16 a.m. EDT June 12, 2013



SHARE 235 CONNECT | 84 TWEET | 47 COMMENT | EMAIL | MORE

Edward Snowden's ability to extract sensitive data from the NSA, working as a low-level contract consultant, comes as no surprise to the security community.

<http://www.usatoday.com/story/news/nation/2013/06/11/snowden-nsa-hacking-privileged-accounts/2412507/>

As Snowden told *The Guardian* in a videotaped interview: "When you're in positions of **privileged access**, like a **systems administrator**, for these sort of high-level government agencies, you're **exposed to a lot more information** on a broad scale than the average employee ... Anybody in the positions of access with that I had could, you know, suck out secrets."

Security experts say Snowden, a Booz Allen Hamilton network analyst based in Hawaii, had the technical savvy to take full advantage of two major security challenges all organizations face: **managing privileged accounts** and keeping PCs, databases and applications **updated with the latest security patches**.

**SharePoint isn't why Snowden breached NSA -- lax security is**

By **Derrick Wlodarz** | Published 3 weeks ago

<http://betanews.com/2013/08/01/sharepoint-isnt-why-snowden-breached-nsa-lax-security-is/>

The information was first picked up by **The Register** after the General was caught making the admission at a recently broadcast cyber security forum (which can be viewed on **YouTube** in its entirety). He described of Snowden: "This leaker was a **sysadmin** who was trusted with moving the information to actually make sure that the right information was on the **SharePoint** servers that NSA Hawaii needed".

The same study found that **65% of respondents** admitted their organizations are **not marking data into categories or sensitivity levels in any way**. And again, while the study has no

U.S. NEWS | Updated June 9, 2013, 11:47 p.m. ET

**Technology Emboldened the NSA**  
*Advances in Computer, Software Paved Way for Government's Data*

[http://online.wsj.com/article/SB100014241278873234956045785352906274429.html?mod=rss\\_US\\_News](http://online.wsj.com/article/SB100014241278873234956045785352906274429.html?mod=rss_US_News)

The NSA also became an early adopter. At a 2009 conference on so-called cloud computing, an NSA official said the agency was developing a new system by linking its various **databases** and using **Hadoop** software to analyze them, according to comments reported by the trade publication **InformationWeek**.

As it has gathered ever more data, the government has had to develop new ways to include **privacy protections** by reworking legal theories and harnessing the same type of data-analysis technology to **monitor how the information is used**, said officials familiar with the programs.




# Regulation ≠ Security





# Technology trends

**Cloud**



private	public	SaaS
---------	--------	------

Data is...

- ✓ Leaving the Data Center
- ✓ Stored on shared drives
- ✓ Hosted by 3<sup>rd</sup> party
- ✓ Managed by 3<sup>rd</sup> party

**Consumerization of IT**

**Mobile**



BYOD	Apps	Social
------	------	--------

Data is...

- ✓ Generated 24x7
- ✓ Used Everywhere
- ✓ Always Accessible
- ✓ On private devices

**Everything is Everywhere**

**BigData**



Hadoop	No-SQL	Files
--------	--------	-------

Data is...

- ✓ Produced in high volumes
- ✓ Stored unstructured
- ✓ Analyzed faster/cheaper
- ✓ Monetized

**Data Explosion**





# Embracing the Trends in Digital Government

- Enable access to info and services anywhere, anytime, on any device
- Procure and manage devices, applications, and *data* in smart, secure and affordable ways [italics mine] – build a sound governance strategy
- Unlock the power of *government data* to spur innovation and improve services for American people [italics mine]



I want us to ask ourselves every day, how are we using technology to make a real difference in people's lives."  
– President Barack Obama

*Digital Government: Building a 21<sup>st</sup> century platform to better serve the American people*

<http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>



# Australia Open Data Initiative

- More efficient and effective government
- Economic development and innovation
- Transparency & accountability

Information courtesy of Pia Waugh Director of Coordination and Gov 2.0  
Office of the Australian Government CTO



**Australian Government**  
**Department of Finance**



**The Coalition's Policy  
for E-Government and  
the Digital Economy**

September 2013

"The Principles on open public sector information form part of a core vision for government information management in Australia. They rest on the democratic premise that public sector information is a national resource that should be available for community access and use.

Office of the Australian Information Commissioner,  
Principles on open public sector information

<http://www.oaic.gov.au/information-policy/information-policy-resources/information-policy-agency-resources/principles-on-open-public-sector-information>





## Australia Public Service Big Data Strategy

### Vision

*The Australian Government will use big data analytics to enhance services, deliver new services and provide better policy advice, while incorporating best practice privacy protections and leveraging existing ICT investments.*

*The Australian Government will be a world leader in the use of big data analytics to drive efficiency, collaboration and innovation in the public sector.*

<http://www.finance.gov.au/sites/default/files/Big%20Data%20Strategy.pdf>



# Some Government Big Data Use Cases



- National defense and security
- Threat prediction and prevention
- Social program fraud, waste and errors
- Tax compliance - fraud and abuse
- Crime prediction and prevention
- Safety and health

Police departments analyze images from aerial cameras, news feeds, social networks to detect events, persons or items of interest



Tax agencies identify fraudsters and support investigation by analyzing complex identity information and tax returns

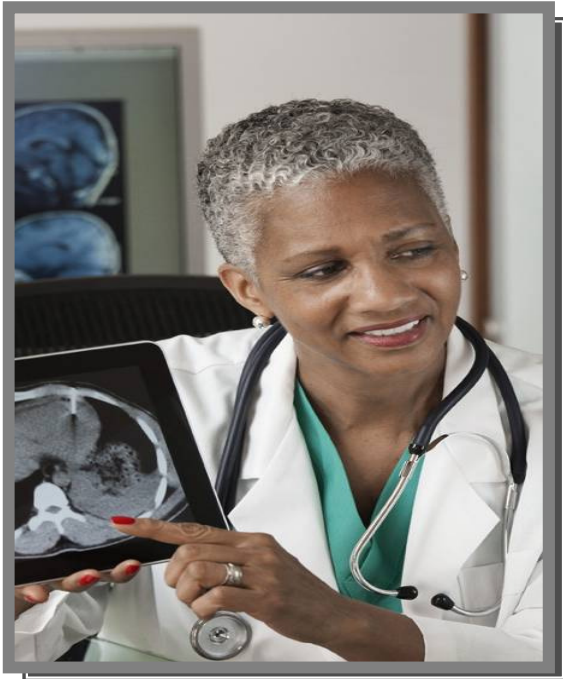


Social program agencies gain a clearer understanding of beneficiaries and proper payments





# Sensitive Data is Common in Big Data Projects



Healthcare



Customer



Citizen



# Sensitive Data Is at Risk



**70%**

of organizations surveyed use live customer data in non-production environments (testing, Q/A, development)

Database Trends and Applications. *Ensuring Protection for Sensitive Test Data*

**\$188**

per record

cost of a data breach

The Ponemon Institute. *2013 Cost of Data Breach Study*

**\$5.4M**

Average cost of a data breach

The Ponemon Institute. *2013 Cost of Data Breach Study*

**50%**

of organizations surveyed have no way of knowing if data used in test was compromised

The Ponemon Institute. *The Insecurity of Test Data: The Unseen Crisis*

**52%**

of surveyed organizations outsource development

The Ponemon Institute. *The Insecurity of Test Data: The Unseen Crisis*



# Assessing the Requirements in Big Data



**The same risks are magnified...**

**...and big data introduces new challenges**

## Data Breach



- Greater attack surface and single repository

## Reputation and trust



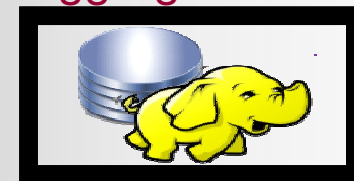
- Greater risk of litigation and fines

## New Users/Apps



- Data sharing and new user access

## “Schemaless” aggregation



- Instant PII

## Compliance



- Changing and new privacy legislation

## BIG DATA PLATFORM



## Fewer Tools



- Traditional tools may not apply





# Security & Privacy Principles for 21<sup>st</sup> Century (US)

"To support information sharing and collaboration, we must build in security, privacy, and data protection throughout the entire technology life cycle."

"We must also adopt new solutions in areas such as continuous monitoring, identity authentication, and access management, and cryptography that support the shift from securing devices to securing the data itself and ensure that data is only shared with authorized users."

"The Federal Government will seek to foster trust, accountability, and transparency about how user information is collected, used, shared, and secured."

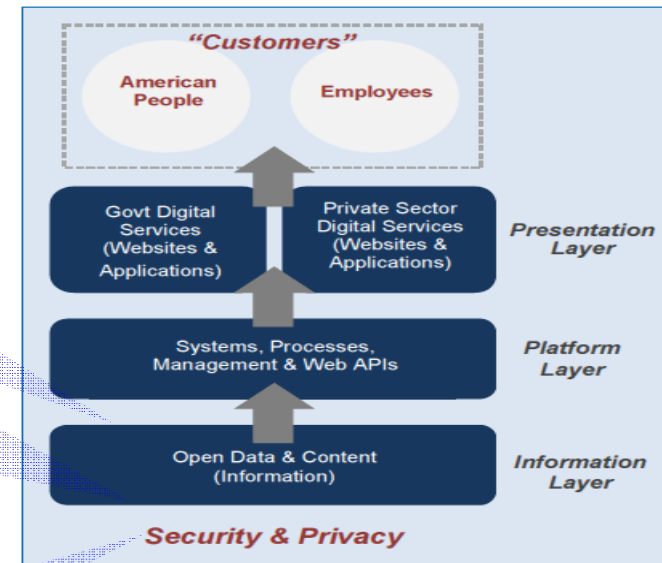


Figure 1: The Layers of Digital Services

**Digital Government: Building a 21<sup>st</sup> century platform to better serve the American people**  
<http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>





## Data privacy and security for big data (Australia)

- Privacy
  - Better practice in linking together cross-agency data sets
  - De-identification
  - Considerations before releasing open data
  - Cross-border flows
- Security
  - Security concerns are more complex
- Data Management and retention
  - Over-retention can be security and privacy risk



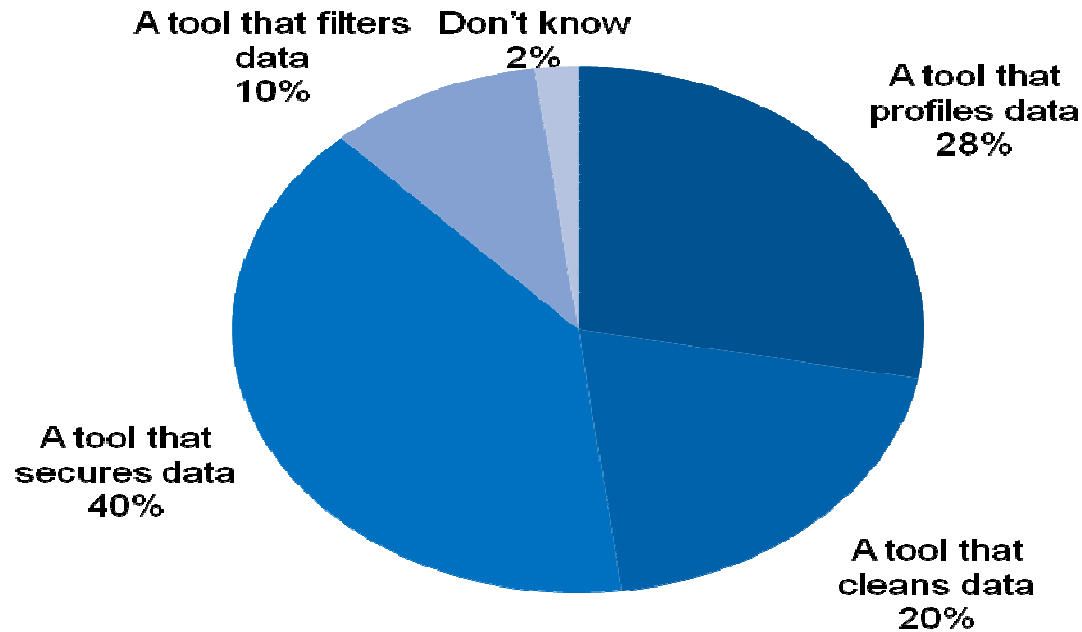
AGIMO is part of the Department of Finance and Deregulation

<http://www.finance.gov.au/sites/default/files/Big%20Data%20Strategy.pdf>



# Security and Privacy: Essential in Big Data

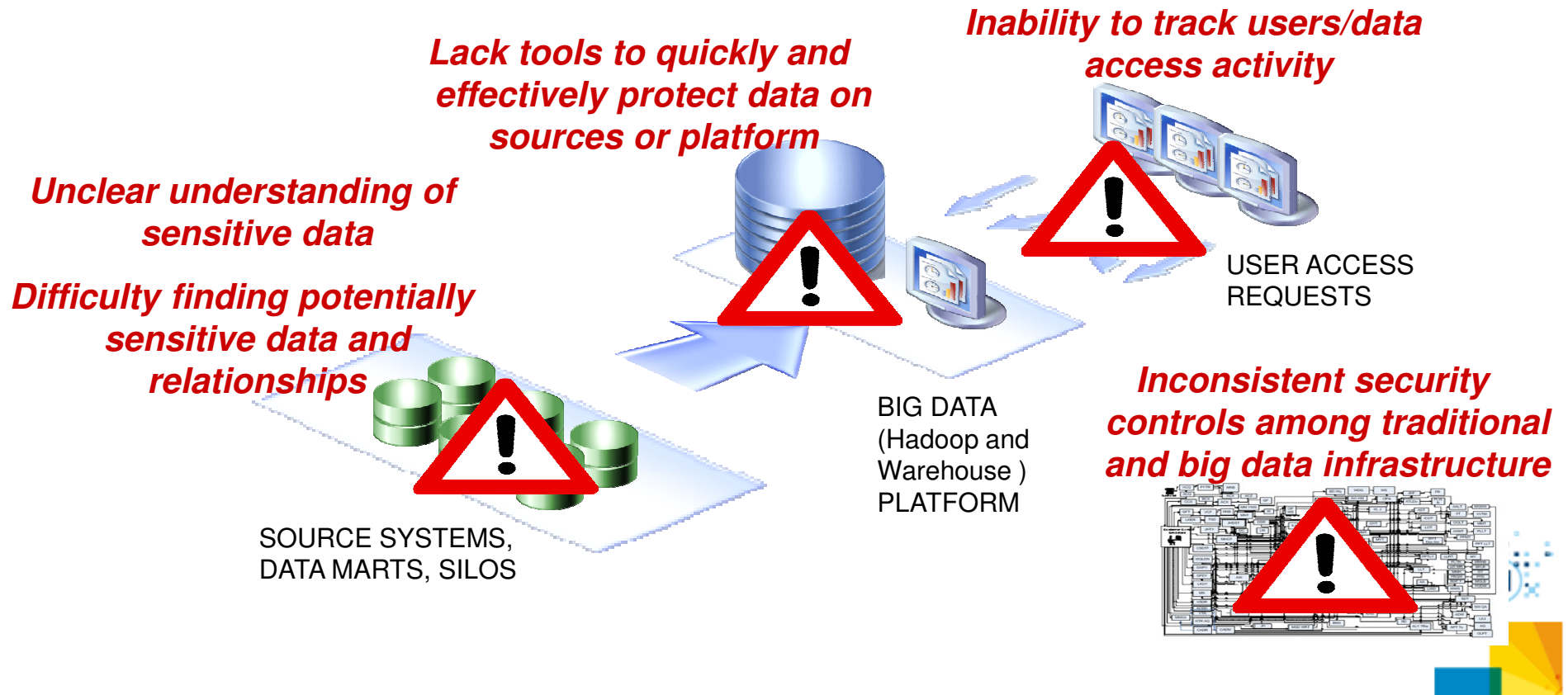
**Q18. Which information governance tool would be the highest investment priority?**



Base: 512 Director or VP level professionals with decision making authority for Big Data technologies  
Source: "IBM Data Governance", a commissioned study conducted by Forrester Consulting on behalf of IBM, July, 2013



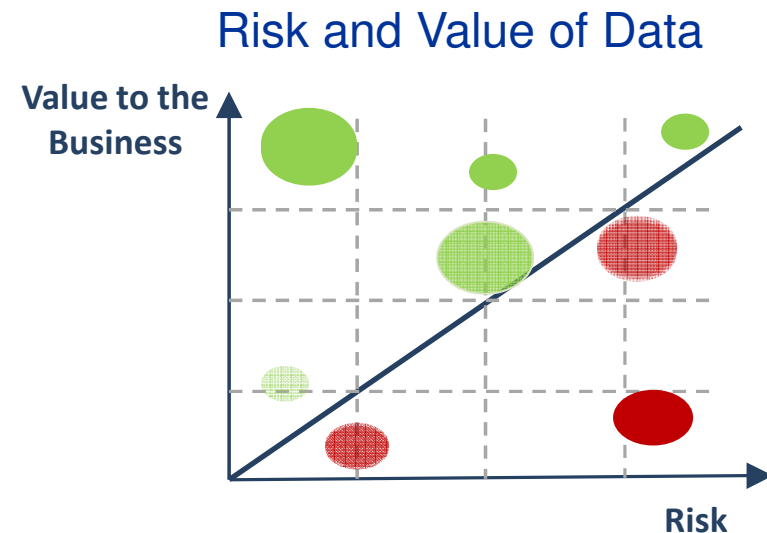
# Technology Barriers to Security and Privacy



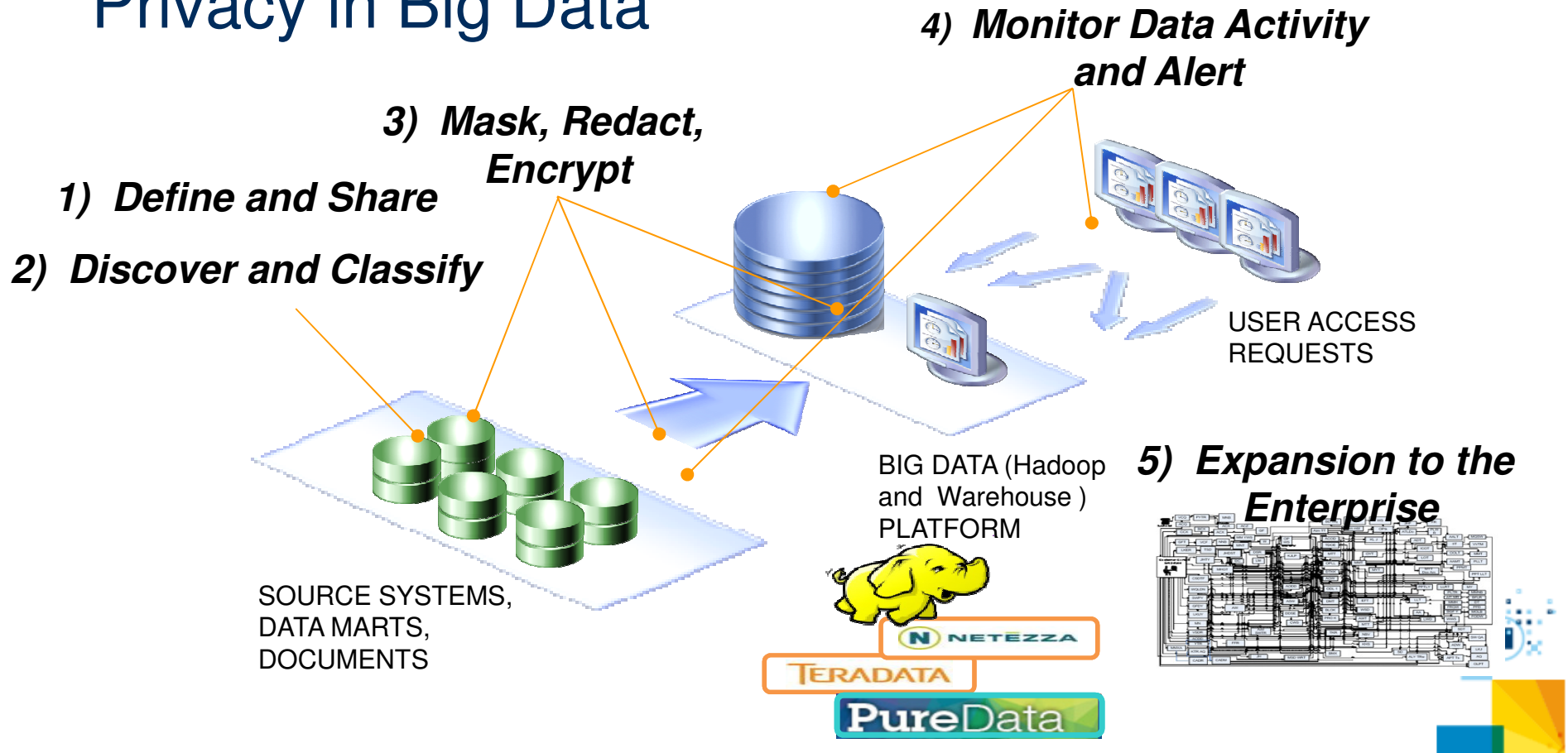


# IBM's Approach to Privacy in Big Data Initiatives

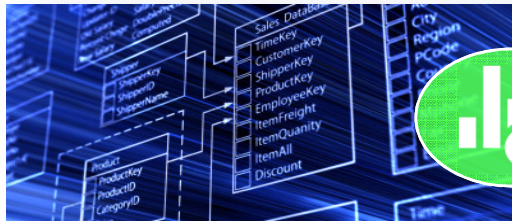
- Understanding what data you have, defining a common vocabulary and assigning risk
  - Define and Share: Business and IT agree on relative data risk, value
  - Discover and Classify: Exploring data sources and plotting the sources
- Mitigating Risk with Data Protection
  - Mask and Redact: Moving the risk areas above the line
- Maintaining a Tolerant Risk Level
  - Monitor Data Activity: Keeping Risk-prone areas above the line
- Expansion to the Enterprise
  - A solution that is extensible across the enterprise



# IBM's Approach to Data Security and Privacy in Big Data



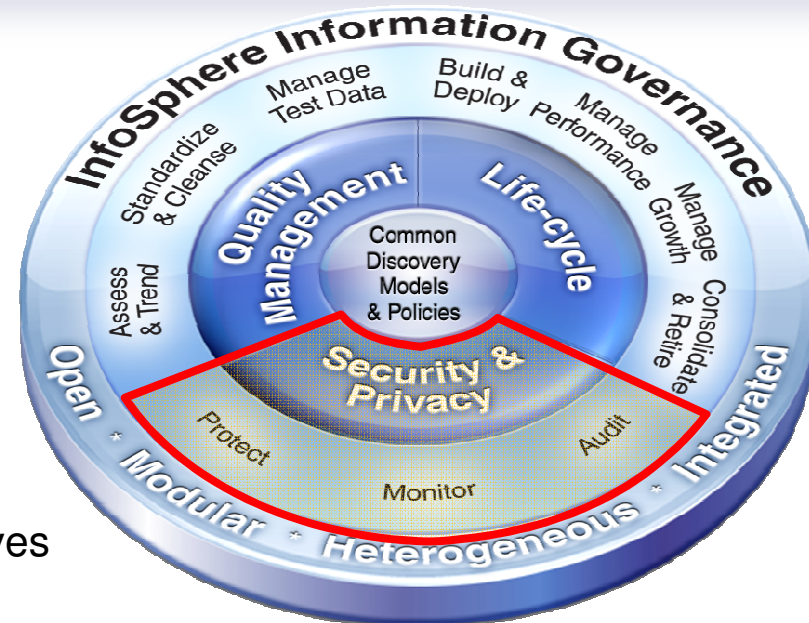
# Information Governance for All Data



Manage, improve and leverage information to increase confidence in enterprise analytics and decisions.

## Challenges

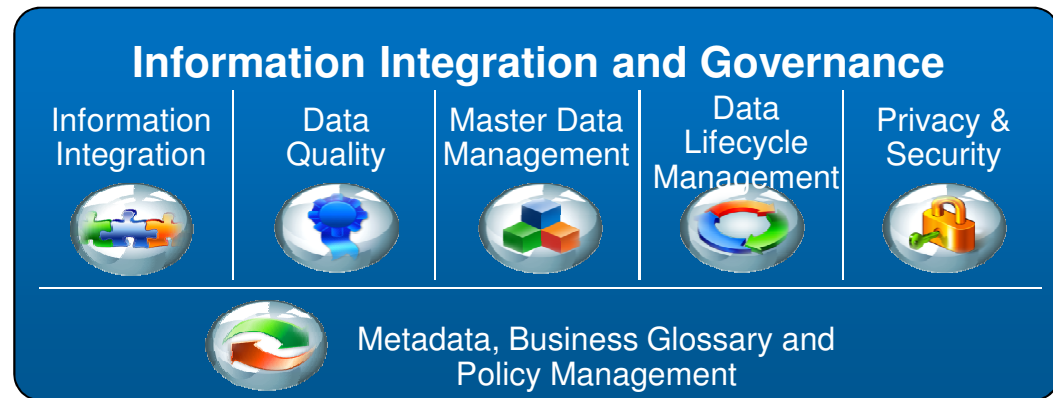
- Controlling data management costs
- Managing legal and regulatory risks
- Maintaining data security and privacy
- Protecting personally identifiable information
- Assessing compliance with business objectives



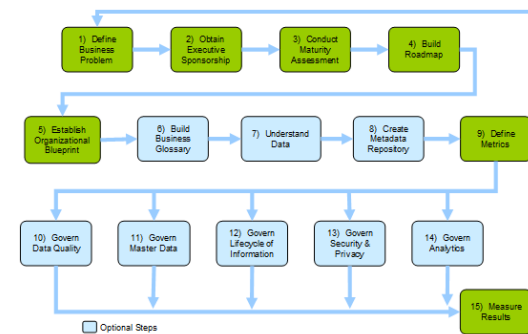
# Governance is More than Technology



- 1. Leadership:**  
IBM founded Information Governance Community in '06
- 2. Assets:**  
Getting started, case studies, metrics and analyst reports
- 3. IBM Unified Process:**  
Best practices from our diverse customer base
- 4. Community:**  
Peer network of information governance champions



## IBM Information Governance Unified Process





# Government of South Australia



## Implementing IBM® InfoSphere® Guardium to protect sensitive information

### The Need:

The Department of Planning, Transport and Infrastructure (DPTI) within the Government of South Australia maintained a large volume of personal and financial information within its driver's license and registration system, which was subject to specific security-compliance requirements. They sought security-management software to better monitor the system and protect personal data.

### The Solution:

IBM InfoSphere Guardium Database Activity Monitor software was implemented within the DPTI driver's license and vehicle registration system. The software would track user activity, local changes to the database structure and the database schema for improper or fraudulent use. The solution also monitored all user activity and acted as a single, read-only repository for all audit data.

### The Benefits:

- Effectively monitor Privileged-user activity
- Control data viewed by users and administrators
- Effectively manage local activity on a monitored server

### Solution Components:

- IBM®InfoSphere®  
Guardium Database  
Activity Monitor



Government of South Australia  
Department of Planning,  
Transport and Infrastructure



# Asian Government Agency



Implementing IBM Information Integration and Governance to streamline collection and maintenance of citizen card data

## The Need:

To maintain accurate data for its citizen cards, this agency must gather information from multiple departments within the government. The data was silo'd and difficult to gather in a timely manner. Moving forward, the agency needed a solution that would help it integrate with other departments.

## The Solution:

IBM InfoSphere IIG software provides a platform to collect, maintain, analyze, and secure accurate citizen card information.

## The Benefits:

- Established integrated platform to streamline collection and maintenance of citizen data
- A single view of information enable the agency to analyze trends and assign resources appropriately
- Guardium software helps protect and maintain the integrity of the citizen data

## Solution Components:

- IBM InfoSphere Data Replication
- IBM InfoSphere DataStage
- IBM InfoSphere Master Data Management
- IBM InfoSphere Guardium Database Activity Monitor





# How do you get started with Big Data Security and Privacy?

## Get Educated

- Download educational pieces:  
[Top Tips for Securing Big Data eBook](#)  
[Planning a Hadoop Data Security Deployment](#)
- See what the analysts are saying: [Big Data Needs Agile Governance](#)
- Visit the [InfoSphere Data Privacy for Hadoop page](#)
- Checkout a [video chat](#) between IBM's CPO and the Exec Director of the Future of Privacy Forum

## Schedule a Client Value Engagement (CVE)

- Business and IT: Narrow the communication gap
- Easy to follow programmatic client-centric approach – determine possible benefits from solution
- Fast time to completion: Less than 2 weeks – deliverables easy to follow and understand





## Three Key Imperatives for Big Data & Analytics Success

Build a culture that  
infuses analytics  
everywhere

**Imagine It.**

Invest in a  
big data & analytics  
platform

**Realize It.**

Be confident with  
privacy, security  
and governance

**Trust It.**





# Thank you

**Big Data & Analytics**





# Data Privacy Regulation: A View from the US

- US Govt- Data Privacy Act of 1974
- No unified data privacy regulation
  - Approach based on legislation, regulation and self regulation.
  - By sector (Fair credit reporting act (1970, amended in 2003 for identity theft protection) , Health Information Portability and Accountability Act (HIPAA) 1996 and HITECH (2009)
  - By state (For example California SB 1386 expands on privacy law to include breach notification requirements. )
- US companies required to register with Safe Harbor when working with EU



# CIO Recommendations

CIO Council Releases New Recommendations on Standardized Digital Privacy Controls

- Three key privacy controls:
  - PII Inventory (current and future usage)
  - Privacy impact assessment
    - Consider risk that data *can be combined* with other data...
  - Privacy notice

Federal agencies must identify and address privacy issues and risks at the earliest stages of developing digital programs and services--*well before* data about individuals are collected.. (p. 4)

Recommendations for Standardized Implementation of Digital Privacy Controls, December 2012  
Product of the Federal Chief Information Officers Council  
[https://cio.gov/wp-content/uploads/downloads/2012/12/Standardized\\_Digital\\_Privacy\\_Controls.pdf](https://cio.gov/wp-content/uploads/downloads/2012/12/Standardized_Digital_Privacy_Controls.pdf)



# States: Security and privacy voted #1 Strategic priority for 2014



**1. Security:** risk assessment, governance, budget and resource requirements, security frameworks, data protection, training and awareness, insider threats, third party security practices as outsourcing increases, determining what constitutes “due care” or “reasonable”

2. Consolidation/Optimization
3. Cloud services
4. Project and portfolio management
5. Strategic IT planning
6. Budget and cost control
7. Mobile services / mobility
8. Shared services
9. Interoperable nationwide public safety BB network
10. Health care (ACA)

[http://www.nascio.org/publications/documents/NASCIO\\_StateCIOTop10For2014.pdf](http://www.nascio.org/publications/documents/NASCIO_StateCIOTop10For2014.pdf)

From top 10 Priority Technologies and Tools

1. Cloud
2. Security enhancement tools
3. Mobile







## APP – “Reasonable steps”

- Governance
- ICT security
- Data breaches
- Physical security
- Training
- Workplace policies
- The information lifecycle
- Standards
- Regular monitoring and review



**Australian Government**

**Office of the Australian Information Commissioner**

